Non-CG methylation and multiple epigenetic layers associate child abuse with immune and small GTPase dysregulation

Pierre-Eric Lutz[1,†*], Marc-Aurèle Chay[1,*], Alain Pacis[2], Gary G Chen[1], Zahia Aouabed[1], Elisabetta Maffioletti[3], Jean-François Théroux[1], Jean-Christophe Grenier[2], Jennie Yang[1], Maria Aguirre[4], Carl Ernst[1,5], Adriana Redensek[6], Léon C. van Kempen[4], Ipek Yalcin[7], Tony Kwan[6], Naguib Mechawar[1,5], Tomi Pastinen[6,#] and Gustavo Turecki[1,5]

Affiliations:

[1] McGill Group for Suicide Studies, Douglas Mental Health University Institute, McGill University, Montréal, Canada

[†] Current address: Centre National de la Recherche Scientifique, Institut des Neurosciences Cellulaires et Intégratives, Université de Strasbourg, Fédération de Médecine Translationnelle de Strasbourg, Strasbourg, France

[2] Department of Genetics, CHU Sainte-Justine Research Center, Montréal, Canada

[3] Genetics Unit, IRCCS Istituto Centro San Giovanni Fatebenefratelli, Brescia, Italy

[4] Segal Cancer Centre, Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Canada

[5] Department of Psychiatry, McGill University, Montréal, Canada

[6] Department of Human Genetics, McGill University, Montréal, Canada

[#] Current address: Center for Pediatric Genomic Medicine, University of Missouri-Kansas City School of Medicine, Kansas City, USA

[7] Centre National de la Recherche Scientifique, Institut des Neurosciences Cellulaires et Intégratives, Université de Strasbourg, Fédération de Médecine Translationnelle de Strasbourg, Strasbourg, France

*, equal contribution

Supplementary Material

Contents

NB: Supplementary Tables have been deposited and are freely available on the 1st author's ResearchGate webpage (researchgate.net/profile/Pierre_Eric_Lutz).

Online methods

Human samples & tissue dissections

Post-mortem lateral amygdala brain tissue was obtained in collaboration with the Quebec Coroner's Office, from the Douglas-Bell Canada Brain bank (douglasbrainbank.ca/, Montreal, Canada). This study included (i) subjects who died suddenly without prolonged agonal state or protracted medical illness, and with no history of psychiatric disorder (Controls, C, N=17), and (ii) subjects with a history of severe child abuse, who died by suicide in the context of a major depressive episode (Early-life adversity, ELA, N=21). Sample characteristics are presented in Supplementary Table1. Groups were matched for age, post-mortem interval (PMI) and brain pH. Psychological autopsies were performed by trained clinicians on both controls and cases, with the informants best-acquainted with the deceased, as described previously[1] and as validated by our group and others[2-7]. Diagnoses were assigned based on DSM IV criteria. Characterization of early-life histories was based on adapted Childhood Experience of Care and Abuse (CECA) interviews assessing experiences of sexual and physical abuse, as well as neglect[8, 9], and for which scores from siblings are highly concordant[2, 9]. We considered as severe early-life adversity reports of non-random major physical and/or sexual abuse during childhood (up to 15 years). Only cases with the maximum severity ratings of 1 and 2 were included. This information was then complemented with medical charts and coroner records. Ethical approval was obtained from the Institutional Review Board of the Douglas Mental Health University Institute. Written informed consent was obtained from the families of each of the deceased subjects prior to inclusion in the study.

Next-Generation Sequencing

ChIP-seq, WGBS, and RNA-Seq experiments were carried out by expert technicians at the McGill University and Genome Quebec Innovation Center, following standard operating procedures from the International Human Epigenome Consortium (IHEC, see ihec-epigenomes.org/).

Data availability

Sequencing data and source files will be made publicly available during regular data releases on the IHEC consortium data portal, and on NCBI's GEO website. Any additional data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code availability

All analytical tools used in this study (see details in sections below) have already been published and are freely available. These notably include: the BSmooth[10] and DESeq2[11] R

packages for differential DNA methylation and gene expression analysis, respectively; the python scripts region_analysis[12] and deepTools[13]; BWA[14]; and diffReps for differential enrichment of ChIP-Seq reads[12].

ChIP-seq library preparation

Because of the small size of the amygdala lateral nucleus, and the large amounts of tissue required for multiple immune-precipitations and the ChIP-seq analysis of 6 histone marks, tissue from 17 C and 21 ELA subjects were distributed into 7 ELA and 4 C pools, for an average of 472 mg of tissue available for ChIP-Seq experiments per pool (see Supplementary Table2). Libraries were prepared using the automated protocol for the Kapa HTP Library Preparation Kit (Illumina), and sequencing was performed using the Illumina HiSeq2000, as per the manufacturer's instructions, to achieve at least 30 and 60 million reads for narrow (H3K27ac, H3K4me3) and broad (H3K27me3, H3K36me3, H3K4me1, H3K9me3) marks, respectively (FigS1a).

ChIP-seq data processing

Trimmomatic[15], BWA[14], Picard and deepTools[13] were used to pre-process and align the sequencing reads. Global visualization for the ChIP-seq data was accomplished using IGV[16] and ngs.plot[17]. Inter-sample correlations and hierarchical clustering were achieved using deepTools. Identification of differential enrichment sites for each histone mark was done using diffReps with window size 1000bp and sliding step 100bp[12]. A FDR <10% and p <0.0001 for negative binomial test were used as significance cutoffs. ChromHMM was used to partition the genome into 200bp bins, and annotate them to chromatin states[18]. A 10-state model was chosen and applied to all data sets. For each clinical group, a consensus epigenomic map was defined using the genomic regions showing at least a 70% agreement between the samples for a state (at least 3/4 C pools and 5/7 ELA pools). State transitions (ST) were then defined as regions with differing states between the consensus C map and ELA map. For characterization of the distribution of DS and ST, we used the region_analysis package to annotate them to genomic features[12].

WGBS library preparation

Whole-genome sequencing libraries were generated from 700 to 1,000 ng of genomic DNA spiked with 0.1% (w/w) unmethylated λ DNA (Promega) previously fragmented to 300–400 base-pairs (bp) peak sizes using the Covaris focused-ultrasonicator E210. Fragment size was controlled on a Bioanalyzer DNA 1000 Chip (Agilent) and the KAPA High Throughput Library Preparation Kit (KAPA Biosystems) was applied. End repair of the generated dsDNA with 3′- or 5′-overhangs, adenylation of 3′-ends, adaptor ligation and clean-up steps were carried out

as per KAPA Biosystems' recommendations. The cleaned-up ligation product was then analysed on a Bioanalyzer High Sensitivity DNA Chip (Agilent) and quantified by PicoGreen (Life Technologies). Samples were then bisulfite converted using the Epitect Fast DNA Bisulfite Kit (Qiagen), according to the manufacturer's protocol. Bisulfite-converted DNA was quantified using OliGreen (Life Technologies) and, based on quantity, amplified by 9–12 cycles of PCR using the Kapa Hifi Uracil+DNA polymerase (KAPA Biosystems), according to the manufacturer's protocol. The amplified libraries were purified using Ampure Beads and validated on Bioanalyzer High Sensitivity DNA Chips, and quantified by PicoGreen. Libraries were run on an Illumina HiSeq 2000 (100bp paired-end), yielding ≈164 million reads/library on average (Fig.S3c).

WGBS data processing

As previously described[19], in-house generated methylome libraries were aligned using BWA 0.6.1[14] after converting all the reads in bisulfite mode to the human hg19/GRCh37 genome reference. Both reads in a pair were trimmed of any low-quality sequence at their 3' ends (with Phred scale score >=30). Post-process read mappings were made as previously described[19], including clipping 3' ends of overlapping read pairs in both forward and reverse strand mappings, filtering duplicate, low-mapping quality reads, read pairs not mapped at the expected distance based on the library insert size as well as reads with more than 2% mismatches. Methylation calls of individual CpGs were extracted using Samtools in mpileup mode. CGs overlapping SNPs from dbSNPs (137) and CpGs located within ENCODE DAC blacklisted regions or Duke excluded regions[20] were discarded.

Methylation data characterization. All analyses conducted to characterize the genome-wide abundance and distribution of CG and CAC methylation were done by focusing on cytosines showing a coverage >= 5 (Fig.2 and Fig.S4-6). We used the region_analysis package[12] to assign each cytosine to a genomic feature (Fig.S5), using the Ensembl v75 annotation for consistency with RNA-Sequencing data analysis (see below).

Differential methylation analysis. Differential methylation analysis was conducted using BSmooth, as described previously[21]. The context of each C was determined, which allowed us to classify each C of the genome as CG or CAC. Methylation levels for each site were estimated by counting the number of reported C ('methylated' reads) divided by the total number of reported C and T ('methylated' plus 'unmethylated' reads) at the same position of the reference genome. To identify differentially methylated regions in the CG context, we performed a strand-independent analysis of CG methylation where counts from the two Cs in a CG and its reverse complement (position *i* on the plus strand and position *i*+1 on the minus strand) were combined and assigned to the position of the C in the plus strand. The summarized methylation estimates of strand-merged CG sites from the 21 ELA and 17 control

samples were used to identify differences in methylation, using the R package BSmooth/BSseq[10] at default parameters. To minimize noise in methylation estimates due to low-coverage data, we restricted the differential methylation analysis to CpG sites with coverage of ≥4 sequence reads in at least 10 samples in each condition, which still allowed us to interrogate changes in methylation levels at ~18 million CG and ~39 CAC million sites. The same strategy was applied for differential methylation analysis in the CAC context, except that by definition methylation data originated for each CAC site from one DNA strand only. We identified differentially methylated regions (DMRs) as regions containing at least 5 consecutive CG, or CAC, sites that were significantly differentially methylated using an unpaired Welch t-test ($p < 0.001$) and that exhibited at least a 1% difference in mean methylation levels between ELA and C groups. Genomic features were attributed to DMRs using the region_analysis package, similar to the annotation of ChIP-Seq DS or ST.

RNA-Sequencing library preparation

RNA was extracted from homogenized brain samples using the RNeasy Lipid Tissue Mini Kit (Qiagen). Quantity and quality of extracted RNAs were measured using an Agilent 2100 Bioanalyzer. RNA-Sequencing libraries were prepared by expert technicians at the McGill University and Genome Quebec Innovation Center, using IHEC procedures. Briefly, we used the TrueSeq Stranded Total RNA Sample Preparation kit (Illumina), using the Ribo-Zero Gold kit (Illumina) for depletion of ribosomal RNA, followed by first and second strand cDNA synthesis and fragmentation of dsDNA. Then, fragmented DNA was used for A-tailing, adaptor ligation and 12 cycles of PCR amplification. Libraries were quantified using high sensitivity chip on a Labchip (PerkinElmer), quantitative PCR (KAPA Library Quantification, Kapa Biosystems), and PicoGreen (Life Technologies). Three libraries were run per lane of an Illumina HiSeq 2000 (100bp paired-end), yielding ≈54 million reads/library (Fig.S13b).

RNA-Sequencing data processing

Alignment, counting, and differential expression analysis. As described previously[22, 23], we used: FASTX-Toolkit (hannonlab.cshl.edu/fastx_toolkit/links.html) and Trimmomatic[15] for adapter trimming; Bowtie2 for alignment ; TopHat[24] for transcript alignment; HTSeq-count or Kallisto for counting ; and DESeq2[11] for differential expression analysis. *Alignment*. Following high-throughput sequencing, 100bp paired-end reads were aligned to the hg19 human genome using TopHat v2.1.0 (tophat.cbcb.umd.edu/) with a mate insert distance of 75 bp (-r) and library type fr-unstranded. Reads passing a mapping quality of at least 50 were used for gene and transcript quantification. *Quantification*. Gene annotations from the Ensembl release 75 were used for gene-level quantification quantification. First, we used HTSeq-count version 0.6.1p1 (www-huber.embl.de/users/anders/HTSeq/doc/overview.html)[25], using the

intersection-nonempty mode, and results were combined to form a count matrix of 20,893 transcribed RNAs across 50 samples. As an alternative strategy to HTSeq-count, we also processed reads through the pseudo-aligner Kallisto[26]. Here, expression counts were obtained for isoforms using Kallisto (v. 0.43.0). Then, the tximport (v. 1.0.3) R package was used to reconstruct gene-level counts using the isoform-level counts generated by Kallisto. *Differential expression analysis*. Genes with no mapped fragments were removed from the analysis. Furthermore, genes with low counts were removed by keeping only those with at least 20 counts per subject in average. For both HTSeq-count or Kallisto gene-level counts (see Fig.S14a), differential expression analysis was performed using the DESeq2 general linear model (GLM) using the following covariates: gender[27], age[28], and RIN[29], based on previous literature documenting their impact on human brain RNA-Seq datasets.

Gene Set Enrichment Analysis (GSEA). GSEA was performed as previously described[23, 30]. Log_2 fold changes were obtained for each gene from the differential gene expression analysis. Genes were ranked based on their fold changes where genes with the highest positive fold changes were at the top of the list and those with the lowest negative fold changes were at the bottom of the list. The ranked gene list was then used as an input for the GSEAPreranked tool, with the "classic" enrichment score calculation option selected. The C2 curated gene sets molecular signatures database was used to identify enriched gene sets.

Gene ontology

We used the GREAT tool to identify the enrichment of gene categories in differential sites (DS) or state transition sites (ST) obtained from ChIP-Seq experiments[31]. DS and ST were associated with genes using the default proximal (5kb upstream, 1kb downstream of TSS) and distal (+/- 1Mb of TSS) definition of regulatory regions. Biological process and molecular function gene categories were kept if they passed both the hypergeometric and binomial tests with a fold enrichment ≥1.5 and FDR Q≤0.1. Significant GO terms with less than 5 genes associated with ST or DS were discarded. To account for the recurrence of terms across multiple combinations of ST, we calculated a co-occurrence score for each GO term, consisting of the sum of the –log10 of the binomial p-value for each ST enriched in this term, as described by Feng et al[32].

Supplementary Figures legends.

Supplementary Figure 1. Quality controls for chromatin immunoprecipitation sequencing (ChIP-Seq). (a) ChIP-Seq libraries were prepared and sequenced by expert technicians at the Genome Québec Innovation Center, in the framework of the International Human Epigenomics Consortium (IHEC). We compared psychiatrically healthy controls (C) and subjects with a history of early-life adversity (ELA), and analyzed 4 'broad' (H3K9me3, H3K27me3, H3K36me3, H3K4me1) and 2 'narrow' (H3K4me3, H3K27ac) marks, for which we aimed at sequencing roughly 60 and 30 million reads per library, respectively, as per the IHEC consortium's guidelines. A two-way ANOVA indicated that there was no difference among C and ELA groups in terms of sequencing depths [$F(1,63)=0.52$; $p=0.47$]. **(b)** Quality controls analyses showed that samples for narrow marks showed greater than 0.8 and 1.05 Relative Strand Cross and Normalized Strand Cross correlations, respectively, thereby meeting ENCODE consensus thresholds for quality control[33]. Values are mean±sem.

Supplementary Figure 2. Comparison of amygdalar histone mark profiles with datasets from inferior temporal lobe (Inf Temp), anterior caudate (Ant Caud), and peripheral blood mononuclear cells (Blood). Data were downloaded from the Roadmap Epigenomics Consortium (ncbi.nlm.nih.gov/geo/roadmap/epigenomics/) and compared to each C and ELA amygdalar sample from the present study, using deepTools' Pearson correlations and unsupervised clustering[34], for each of the 6 histone marks: **(a)** H3K4me1; **(b)** H3K27ac; **(c)** H3K4me3; **(d)** H3K36me3; **(e)** H3K9me3; **(f)** H3K27me3. Accession numbers: 1) for inferior temporal lobe: datasets GSM772995 (H3K27ac), GSM772993 (H3K27me3), GSM772982 (H3K36me3), GSM772992 (H3K4me1), GSM772996 (H3K4me3), GSM772994 (H3K9me3); 2) for anterior caudate: datasets GSM772832 (H3K27ac), GSM772827 (H3K27me3), GSM772828 (H3K36me3), GSM772830 (H3K4me1), GSM772829 (H3K4me3), GSM772831 (H3K9me3); 3) for blood mononuclear cells: datasets GSM1127145 (H3K27ac), GSM1127130 (H3K27me3), GSM1127131 (H3K36me3), GSM1127143 (H3K4me1), GSM1127126 (H3K4me3), GSM1127133 (H3K9me3). Of note, H3K4me1, H3K27ac and H3K27me3 better discriminated between tissue types than the 3 other marks.

Supplementary Figure 3. Quality controls for whole-genome bisulfite sequencing libraries (WGBS). To compare DNA methylation patterns among psychiatrically healthy individuals (controls, C) and subjects with a history of severe child abuse (early-life adversity, ELA), WGBS libraries were prepared. **(a)** Bisulfite conversion efficiencies were measured for each DNA sample using spiked-in unmethylated lambda DNA, and were similar between groups: C: 99.3±0.07%, ELA: 99.2±0.06% ($t=0.63$, $p=0.53$). **(b)** In addition, over-conversion (i.e. methylated cytosines converted to uraciles during bisulfite conversion) was determined

experimentally using spiked-in fully methylated pUC19 DNA, and similar values were observed between groups: C: 5.6±0.10%, ELA: 5.7±0.05% (t=0.19, p=0.85). **(c)** Each WGBS library was then sequenced on 1 lane of a HiSeq 2000 (100 base pair, paired-end sequencing) at the Génome Québec Innovation Center, yielding similar sequencing depth across groups: C: 164±3 million reads, ELA: 163±3 million reads (t=0.15, p=0.88). **(d)** During the processing of raw sequencing data, duplicates were removed from downstream analysis, and results indicated similar diversity (t=1.18, p=0.25) among libraries from the C (C:22.5±1.2%) and ELA (C:20.8±0.9%) groups. **(e)** The graphs depicts the number of CG sites that met distinct average coverages among the 38 WGBS libraries. For the characterization of genome-wide abundance and distribution of CG and CAC methylation levels (Fig.2 and Fig.S4-6), we focused on cytosines with a coverage >= 5. Cov., coverage. Values are mean±sem.

Supplementary Figure 4. Distribution of non-CG DNA methylation in the human brain lateral amygdala. (a) The graph, taken from Mo et al[35], depicts the distribution in the mouse of non-CG methylation, measured in 3 neuronal subtypes: glutamatergic neurons (Exc), and parvalbumin-expressing (PV) and vasoactive intestinal peptide (VIP)-expressing inhibitory neurons. **(b)** A very similar distribution of genome-wide average non-CG methylation levels was observed in the human brain lateral amygdala, in healthy controls (C) and subjects with a history of child abuse (early-life adversity, ELA). Values are mean±sem. **(c)** While most cytosines in non-CG contexts were unmethylated, a minority of these positions nevertheless showed methylation levels between 5 and 25%, whatever the 3-letter context considered, with a peak between 15 to 20%. These results indicate that while different numbers of cytosines might be methylated across various non-CG contexts, the abundance of DNA methylation (i.e., the proportion of cells affected) at those sites seems relatively homogeneous. Box plots show median and interquartile range, with whiskers representing minimum and maximum values.

Supplementary Figure 5. Distribution of CG and CAC DNA methylation among distinct chromosomes and genomic features. (a) CAC and CG methylation levels were computed across distinct genomic features defined using the region_analysis package[12]. In the CG context, DNA methylation levels strongly varied as a function of the genomic feature [$F(7,252)=953$; $p<0.0001$], while there was no significant difference [$F(1,36)=0.50$; $p=0.48$] among psychiatrically healthy individuals (C) and subjects with a history of early-life adversity (ELA). Post-hoc comparisons confirmed that, as expected, lowest CG methylation levels were observed in promoter regions, in particular within a 250-base pair distance from the TSS (ProximalPromoter), where methylation levels were significantly lower than in any other gene feature ($p<0.0001$). In the CAC context, similar significant and non-significant effects were

found for genomic features [F(7,252)=917; p<0.0001] and for clinical grouping [F(1,36)=0.03; p=0.85], respectively. In contrast with the CG context, lowest CAC methylation levels were observed in pericentromeric regions (defined by region_analysis as regions located between the boundary of a centromere and the closest gene minus 10kbp of that gene's regulatory region[12]), which showed strongly significant differences with all other features (p<0.0001). **(b)** CAC and CG methylation levels were computed in each chromosome across the whole cohort. As expected, DNA methylation levels were much higher in the CG than in the CAC context (2-way ANOVA; context effect: [F(1,1850)=1238951; p<0.0001]), and methylation abundance strongly varied among chromosomes in each context (chromosome effect: [F(24,1850)=2673; p<0.0001]). As expected also, methylation was extremely low in the mitochondrial genome in both the CG and CAC contexts. Values are mean±sem.

Supplementary Figure 6. Correlations between the expression of genes and DNA methylation levels in their sense or antisense strands, in the CG and CAC contexts. The 1000 most highly (top1000) and 1000 most lowly expressed genes were identified using RNA-Sequencing data (see Fig.S13-14), and compared for abundance of DNA methylation in: **(a,d)** both DNA strands; **(b,e)** the strand where genes are located (sense strand), and **(c,f)** the strand antisense to the one where genes are located. Negative correlations between DNA methylation and gene expression were observed in all cases: **(a)** [F(1,74)=736.1; p<2E-16] **(b)** [F(1,74)=742.8; p<2E-16] **(c)** [F(1,74)=615.1; p<2E-16] **(d)** [F(1,74)=145.3; p<2E-16] **(e)** [F(1,74)=119.2; p<2E-16] **(f)** [F(1,74)=142.6; p<2E-16] (2-way repeated measures ANOVA, main effects of gene category, top1000 versus bottom1000 averaged over 100 bins). Results therefore indicate that gene expression is predicted to the same extent by mCAC on either strand, at least for the coverage achieved in this study.

Supplementary Figure 7. Distributions of histone reads across gene bodies and differential sites (DS). For each histone mark, the figure depicts the distribution of reads (average enrichment of ChIP-seq reads over input) across all gene bodies (All Genes). In addition, and as a control, we also analyzed their distribution at genomic sites where Up- and Down-DS were identified between control (C) and early-life adversity (ELA) groups. Values are mean±sem.

Supplementary Figure 8. Identification of genomic features where histone differential sites (DS) were localized. (a) Localization of DS (identified using diffRep, see *Methods*) among distinct genomic features (defined using region_analysis[12]), for each histone mark. The observed distributions were significantly different across histone marks (df=25, χ^2 =1244, p<0.001). In addition, comparisons between observed and expected (genome-wide distribution

of reads among genomic features in the 2 C and ELA groups combined) distributions showed that, for each type of histone modification, ELA-associated DS were non-randomly located in specific genomic features: H3K27ac (df=1, χ^2 =81.2, p<0.001); H3K27me3 (df=1, χ^2 =42.3, p<0.001); H3K36me3 (df=1, χ^2 =138, p<0.001); H3K4me1 (df=1, χ^2 =287, p<0.001); H3K4me3 (df=1, χ^2 =90.9, p<0.001); H3K9me3 (df=1, χ^2 =22.6, p<0.001). **(b)** Analysis of the directionality of DS showed that ELA more frequently associated with decreases (Down-DS) than increases (Up-DS) in read density, as found for 4 marks: H3K4me1 (df=1, χ^2 =231, p<0.001); H3K4me3 (df=1, χ^2 =73, p<0.001); H3K36me3 (df=1, χ^2 =345, p<0.001); H3K27me3 (df=1, χ^2 =228, p<0.001). DS were equally distributed among Up- and Down-DS for the 2 remaining marks: H3K27ac (df=1, χ^2 =0.13, p=0.19); H3K9me3 (df=1, χ^2 =1.7, p=0.19).

Supplementary Figure 9. Integrin signaling enrichment across multiple histone changes and state transitions. GREAT pathway analysis using MsigDB showed recurrent enrichment of the integrin signaling pathway across six types of state transitions, as well as for H3K27ac down-DS (differential sites). Each analysis passed hypergeometric and binomial testing (fold change ≥ 1.5 and Q ≤ 0.1 for both tests). Negative logarithmic P-value is shown for the binomial test**.** Chromatin states: Str-Trans, strong transcription; Wk-Trans, weak transcription; Str-Enh, strong enhancer; Enh, enhancer.

Supplementary Figure 10. Comparison of main metrics for CG and CAC differentially methylated regions (DMR). DMRs were identified using BSmooth (see *Methods*). **(a)** Compared to CG-DMRs, CAC-DMRs were composed of slightly fewer cytosines (CG: 7.63±0.11; CAC: 7.14±0.08; [t(1,1614.1)=3.63, p=2.9E-04]), and **(b)** were smaller (CG: 321±7 bp; CAC: 245±4 bp; [t(1,1494.9)=9.40, p<2.2E-16]). The amplitude of methylation changes observed in subjects from the ELA group was smaller in the CAC than in the CG context, as shown by **(c)** smaller % changes in methylation levels (CG: 7.75±0.05%; CAC: 4.6±0.03%; [t(1,1338.4)=58.89], p<2.2E-16), and **(d)** smaller areaStat values (a metric measuring the statistical strength of methylation changes among cytosines composing each DMR[10]; CG: 32.3±0.5%; CAC: 28.9±0.4%; [t(1,1575.1)=5.45], p=5.8E-08). Box plots show median and interquartile range, with whiskers representing minimum and maximum values.

Supplementary Figure 11. Comparison of histone mark profiles at CG and CAC DMRs. Average enrichments of ChIP-seq reads over input are shown for each histone mark around each type of DMRs (CG, green; CAC, orange). Histone reads enrichment significantly varied within DMRs and flanking regions (+/- 2 kilobases, kb) for the 6 marks (2-way repeated measures ANOVA, main effects of cytosine position, x-axis; p<0.0001). Also, significant differences in histone reads density (y-axis) among the 2 cytosine contexts were observed for

4 marks (H3K27ac, H3K4me3, H3K27me3, H3K4me1; p<0.05), but not for H3K9me3 (p=0.08) or H3K36me3 (p=0.57). Importantly, significant interactions between histone reads density and cytosine position within DMRs were observed for all marks (p<0.0001). In particular, post-hoc comparisons confirmed that, compared with their flanking regions, CAC-DMRs were significantly enriched for H3K36me3 and depleted in H3K9me3; in contrast, CG-DMRs were characterized by significant enrichments for H3K4me1, H3K4me3, and H3K27ac, and a depletion in H3K36me3 (p<0.0001 for each post-hoc comparison). Values are mean±sem.

Supplementary Figure 12. Enrichment of CG and CAC differentially methylated regions (DMRs) in each chromatin state. The figure depicts the enrichment of each chromatin state (as identified using ChromHMM and the combination of all 6 histone marks, see main text) among genomic regions corresponding to CG- and CAC-DMRs, compared to their relative abundance in the overall human genome. CG-DMRs shown in blue, CAC-DMRs shown in red. Chromatin states: Act-Prom, active promoter; Wk-Prom, weak promoter; Flk-Prom, flanking promoter; Str-Trans, strong transcription; Wk-Trans, weak transcription; Str-Enh, strong enhancer; Enh, enhancer; PcR, polycomb repressed; Heterochr, heterochromatin.

Supplementary Figure 13. RNA-Sequencing quality controls. Total RNAs extracted from lateral amygdala tissue (controls, C, n=17; early-life adversity, ELA, n=21) were used for the preparation of RNA-Sequencing libraries and processed in parallel. **(a)** RNA integrity values (RIN) were not significantly different across RNA samples extracted from C (n=17) and ELA (n=21) subjects (Mann-Whitney U=171; p=0.83). **(b)** The number of reads sequenced in each library was similar across C and ELA groups (t-test t=0.72; p=0.48). **(c-d)** Similarly, there was no significant difference in percentages of duplicates (t=1.25; p=0.22) nor in alignment rate (t=0.51; p=0.62) between the 2 groups. Values are mean±sem.

Supplementary Figure 14. RNA-Sequencing results. **(a)** Very similar results (r=0.82, t=199.69, p< 2.2E-16) were obtained using 2 distinct bio-informatic pipelines for the analysis of RNA-Sequencing data. Raw reads were aligned & counted using either HTSeq-count or Kallisto, followed by the analysis of differential expression between C and ELA groups using DESeq2. **(b)** Gene Set Enrichment Analysis (GSEA) of RNA-Sequencing data. Depicted are the 2 single gene lists that achieved the highest normalized enrichment scores in the c2 MSigDB gene set. As shown in the left panel, a collection of genes related to oligodendrocytes and myelin physiology, which we recently found downregulated in the anterior cingulate cortex of subjects with a history of ELA[23], showed an opposite upregulation in the lateral amygdala (normalized enrichment score=2.91; FWER q-value<0.05), suggesting that opposed transcriptional adaptations might occur as a function of ELA between cortical and subcortical

structures in this glial population. This finding was reinforced by the second best gene list (right panel), which was significantly enriched for upregulated genes in our amygdala data (normalized enrichment score=2.75; FWER q-value<0.05). The later gene collection was previously associated with depression in the middle temporal gyrus[36], and was found enriched in myelin-related genes, suggesting that similar stress-related transcriptomic changes may affect oligodendrocytes among distinct portions of the temporal lobe. **(c)** Identification of Gene ontology (GO) processes most consistently affected by early-life adversity (ELA), as identified by the combined analysis of individual histone marks, chromatin states, DNA methylation, and gene expression (see main text).

References

1. Dumais, A.*, et al.* Risk factors for suicide completion in major depression: a case-control study of impulsive and aggressive behaviors in men. *Am J Psychiatry* **162**, 2116-2124 (2005).
2. Brewin, C.R., Andrews, B. & Gotlib, I.H. Psychopathology and early experience: a reappraisal of retrospective reports. *Psychol Bull* **113**, 82-98 (1993).
3. Conner, K.R., Conwell, Y. & Duberstein, P.R. The validity of proxy-based data in suicide research: a study of patients 50 years of age and older who attempted suicide. II. Life events, social support and suicidal behavior. *Acta Psychiatr Scand* **104**, 452-457 (2001).
4. Hawton, K.*, et al.* The psychological autopsy approach to studying suicide: a review of methodological issues. *J Affect Disord* **50**, 269-276 (1998).
5. Kelly, T.M. & Mann, J.J. Validity of DSM-III-R diagnosis by psychological autopsy: a comparison with clinician ante-mortem diagnosis. *Acta Psychiatr Scand* **94**, 337-343 (1996).
6. McGirr, A.*, et al.* Risk factors for completed suicide in schizophrenia and other chronic psychotic disorders: a case-control study. *Schizophr Res* **84**, 132-143 (2006).
7. Brent, D.A.*, et al.* The validity of diagnoses obtained through the psychological autopsy procedure in adolescent suicide victims: use of family history. *Acta Psychiatr Scand* **87**, 118-122 (1993).
8. Bifulco, A., Brown, G.W. & Harris, T.O. Childhood Experience of Care and Abuse (CECA): a retrospective interview measure. *J Child Psychol Psychiatry* **35**, 1419-1435 (1994).
9. Bifulco, A., Brown, G.W., Lillie, A. & Jarvis, J. Memories of childhood neglect and abuse: corroboration in a series of sisters. *J Child Psychol Psychiatry* **38**, 365-374 (1997).
10. Hansen, K.D., Langmead, B. & Irizarry, R.A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**, R83 (2012).
11. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

12. Shen, L.*, et al.* diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One* **8**, e65598 (2013).
13. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).
14. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
15. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
16. Robinson, J.T.*, et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).
17. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
18. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**, 2478-2492 (2017).
19. Johnson, M.D., Mueller, M., Game, L. & Aitman, T.J. Single nucleotide analysis of cytosine methylation by whole-genome shotgun bisulfite sequencing. *Curr Protoc Mol Biol* **Chapter 21**, Unit21 23 (2012).
20. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
21. Pacis, A.*, et al.* Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome Res* **25**, 1801-1811 (2015).
22. Chen, E.S.*, et al.* Molecular convergence of neurodevelopmental disorders. *Am J Hum Genet* **95**, 490-508 (2014).
23. Lutz, P.E.*, et al.* Association of a History of Child Abuse With Impaired Myelination in the Anterior Cingulate Cortex: Convergent Epigenetic, Transcriptional, and Morphological Evidence. *Am J Psychiatry* **174**, 1185-1194 (2017).
24. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
25. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
26. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).
27. Trabzuni, D.*, et al.* Widespread sex differences in gene expression and splicing in the adult human brain. *Nat Commun* **4**, 2771 (2013).
28. Erraji-Benchekroun, L.*, et al.* Molecular aging in human prefrontal cortex is selective and continuous throughout adult life. *Biol Psychiatry* **57**, 549-558 (2005).

29. Gallego Romero, I., Pai, A.A., Tung, J. & Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol* **12**, 42 (2014).

30. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

31. McLean, C.Y.*, et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).

32. Feng, J.*, et al.* Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol* **15**, R65 (2014).

33. Landt, S.G.*, et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831 (2012).

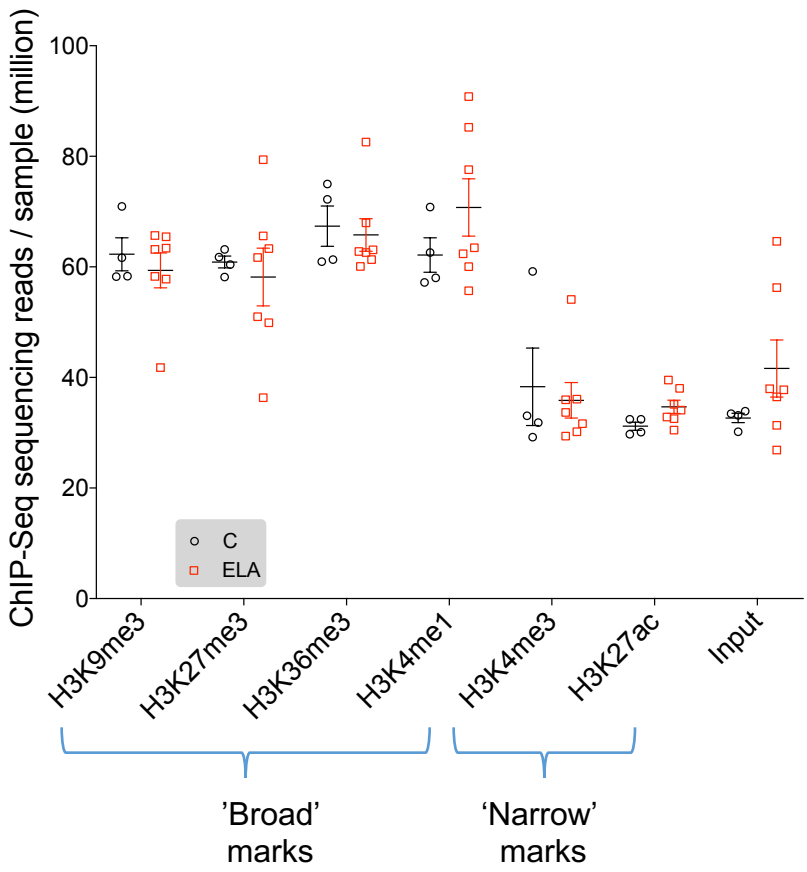
34. Bernstein, B.E.*, et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048 (2010).

35. Mo, A.*, et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369-1384 (2015).

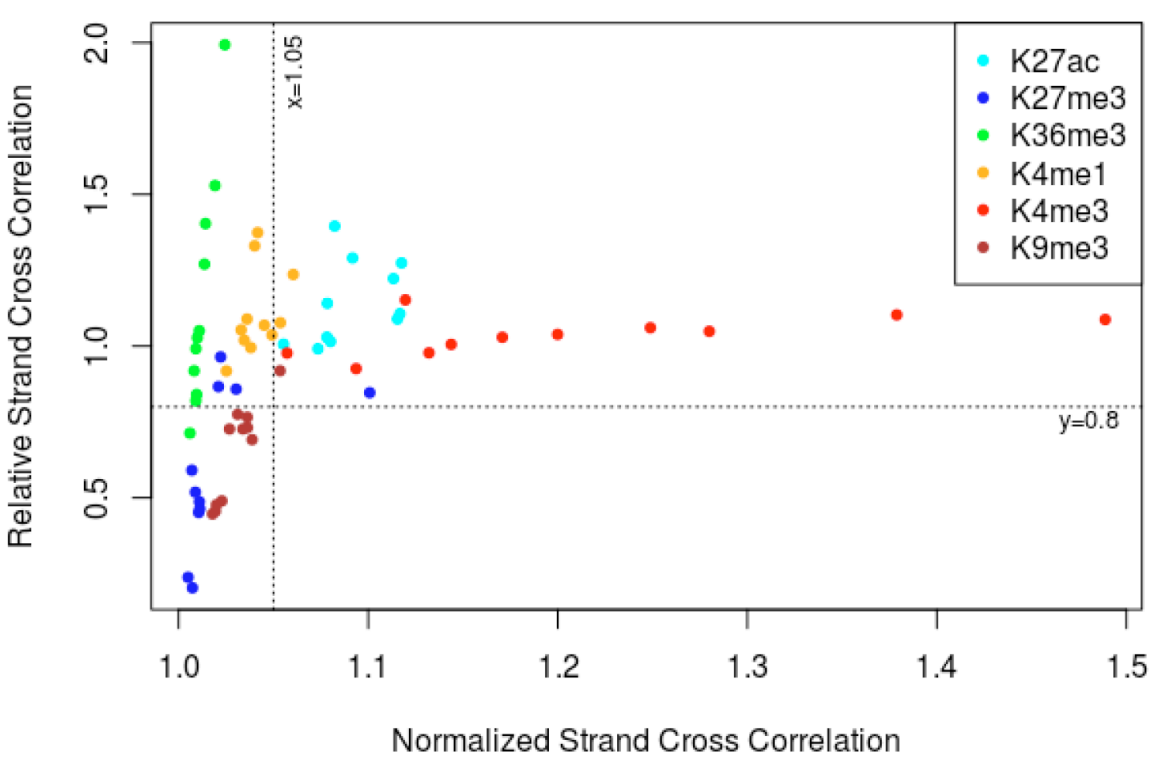
36. Aston, C., Jiang, L. & Sokolov, B.P. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder. *Mol Psychiatry* **10**, 309-322 (2005).

FigureS1

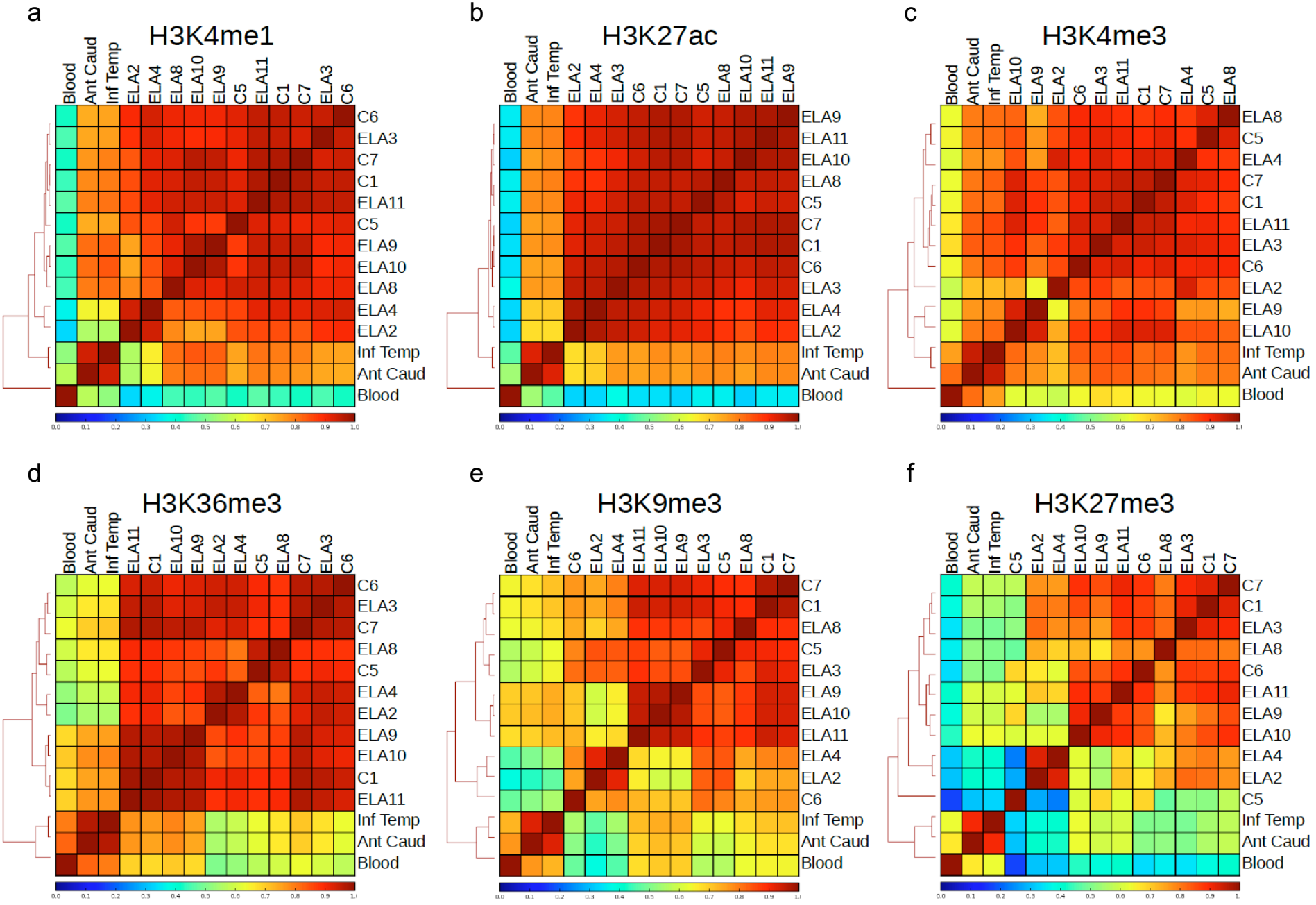
a


ChIP-Seq sequencing reads / sample (million)
100
80
60
40
20
0
C
ELA
H3K9me3
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K27ac
Input
'Broad' marks
'Narrow' marks

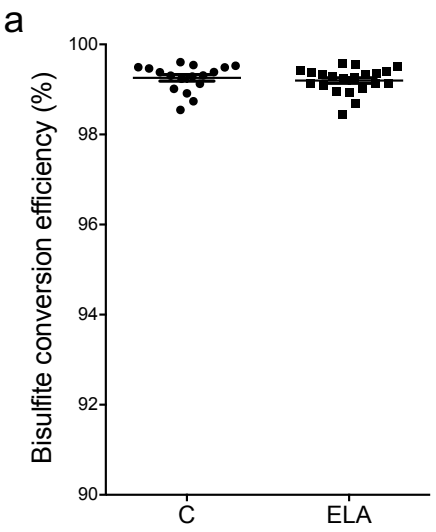
b

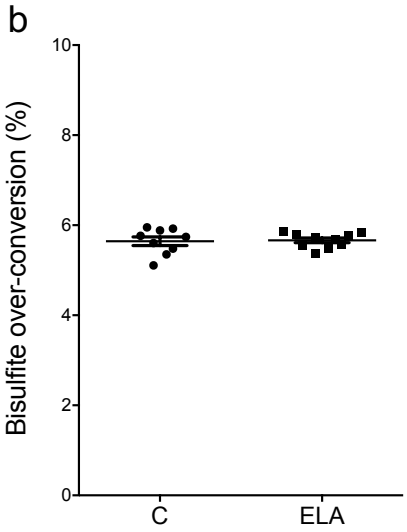

Relative Strand Cross Correlation
2.0
1.5
1.0
0.5
x=1.05
y=0.8
K27ac
K27me3
K36me3
K4me1
K4me3
K9me3
1.0
1.1
1.2
1.3
1.4
1.5
Normalized Strand Cross Correlation

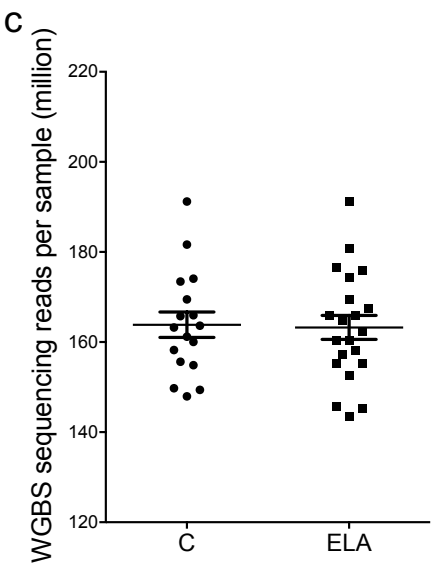
FigureS2

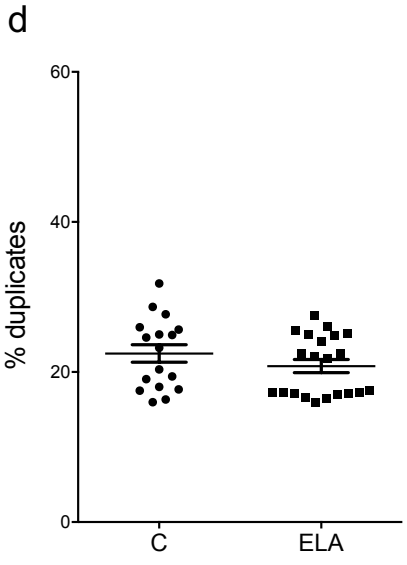

a
H3K4me1
Blood
Ant Caud
Inf Temp
ELA2
ELA4
ELA8
ELA10
ELA9
C5
ELA11
C1
C7
ELA3
C6
C6
ELA3
C7
C1
ELA11
C5
ELA9
ELA10
ELA8
ELA4
ELA2
Inf Temp
Ant Caud
Blood
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
b
H3K27ac
Blood
Ant Caud
Inf Temp
ELA2
ELA4
ELA3
C6
C1
C7
C5
ELA8
ELA10
ELA11
ELA9
ELA9
ELA11
ELA10
ELA8
C5
C7
C1
C6
ELA3
ELA4
ELA2
Inf Temp
Ant Caud
Blood
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
c
H3K4me3
Blood
Ant Caud
Inf Temp
ELA10
ELA9
ELA2
C6
ELA3
ELA11
C1
C7
ELA4
C5
ELA8
ELA8
C5
ELA4
C7
C1
ELA11
ELA3
C6
ELA2
ELA9
ELA10
Inf Temp
Ant Caud
Blood
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
d
H3K36me3
Blood
Ant Caud
Inf Temp
ELA11
C1
ELA10
ELA9
ELA2
ELA4
C5
ELA8
C7
ELA3
C6
C6
ELA3
C7
ELA8
C5
ELA4
ELA2
ELA9
ELA10
C1
ELA11
Inf Temp
Ant Caud
Blood
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
e
H3K9me3
Blood
Ant Caud
Inf Temp
C6
ELA2
ELA4
ELA11
ELA10
ELA9
ELA3
C5
ELA8
C1
C7
C7
C1
ELA8
C5
ELA3
ELA9
ELA10
ELA11
ELA4
ELA2
C6
Inf Temp
Ant Caud
Blood
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
f
H3K27me3
Blood
Ant Caud
Inf Temp
C5
ELA2
ELA4
ELA10
ELA9
ELA11
C6
ELA8
ELA3
C1
C7
C7
C1
ELA3
ELA8
C6
ELA11
ELA9
ELA10
ELA4
ELA2
C5
Inf Temp
Ant Caud
Blood
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

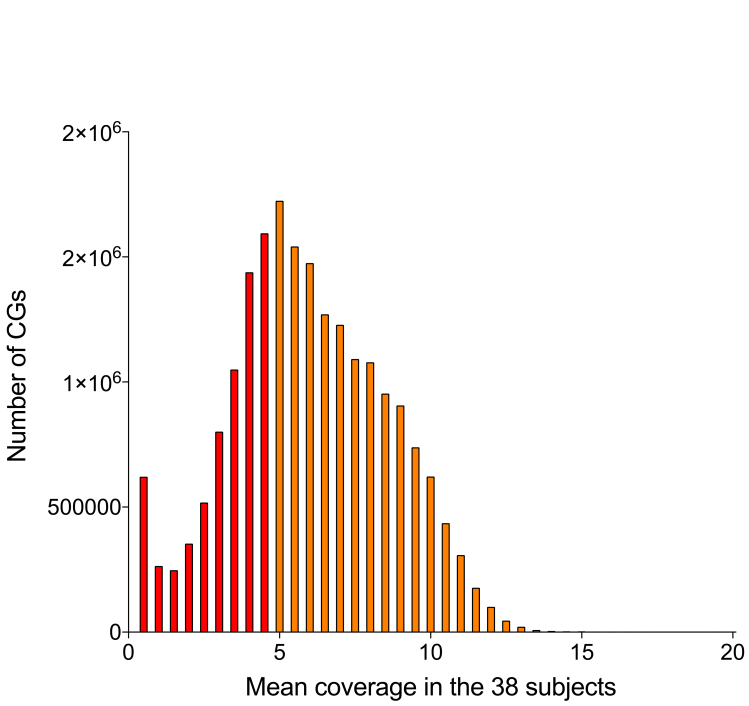
FigureS3











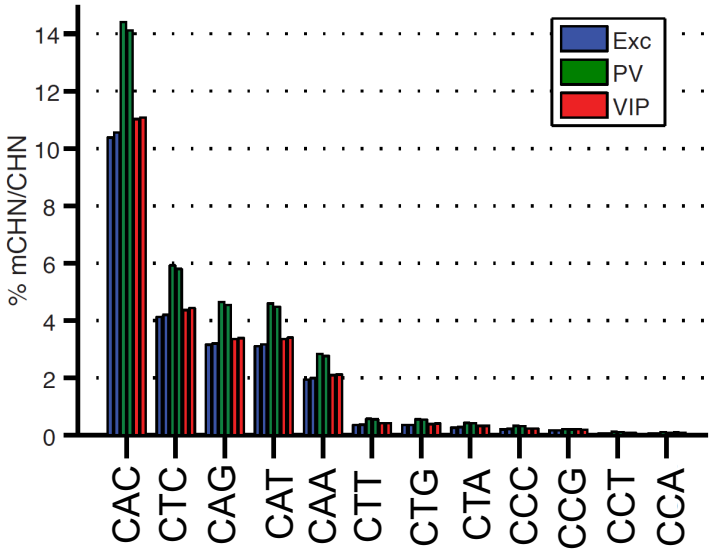
Nbr of CGs with

- Mean cov.≥1 19 957 311
- **Mean cov.≥5 13 704 913**
- Mean cov.≥10 1 715 486

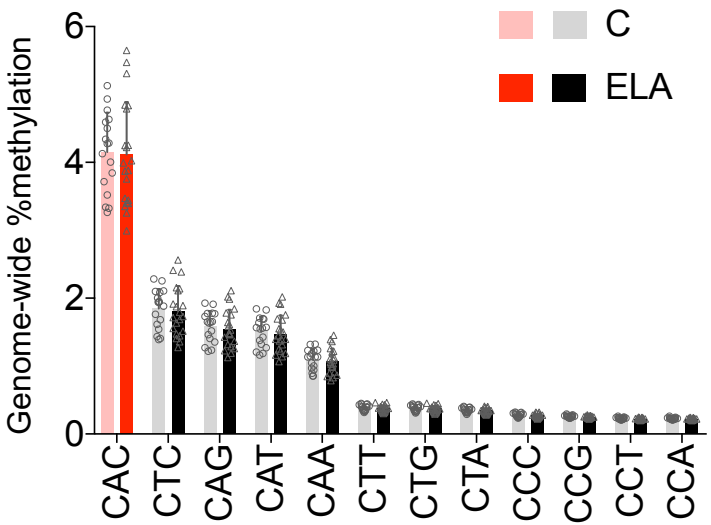
FigureS4

a

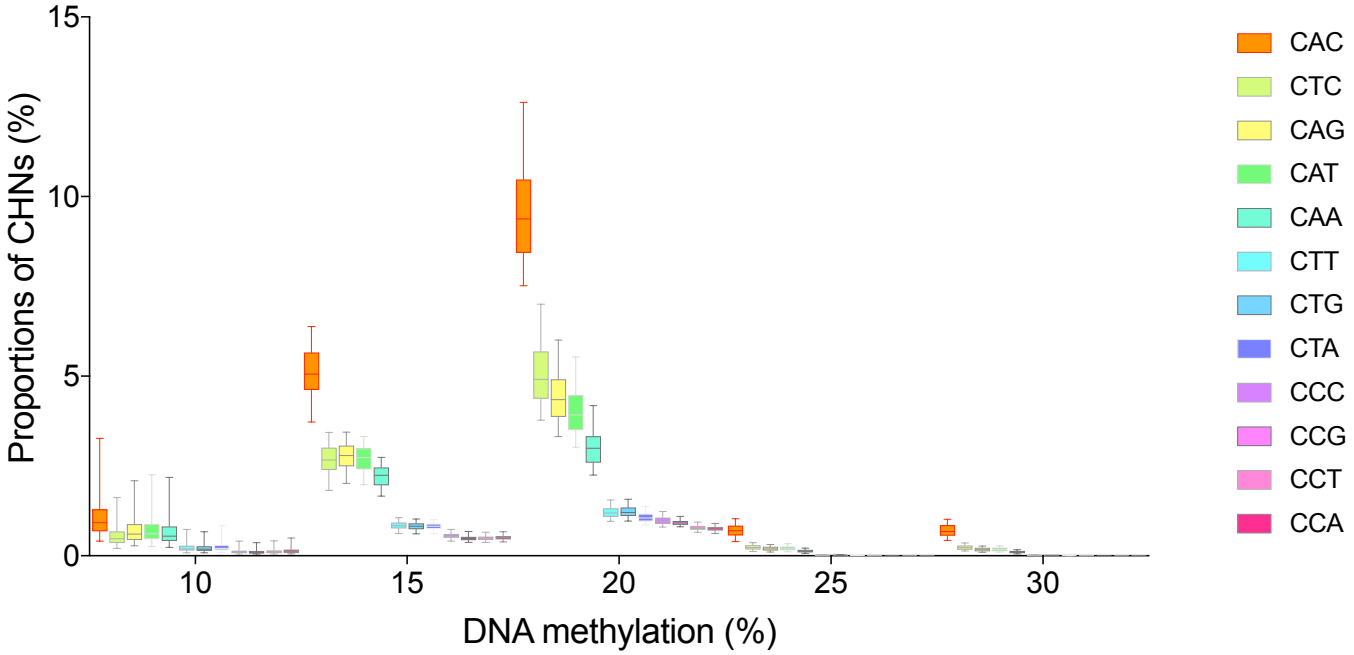
Taken from Mo et al, Neuron 2015; 86(6):1369-84



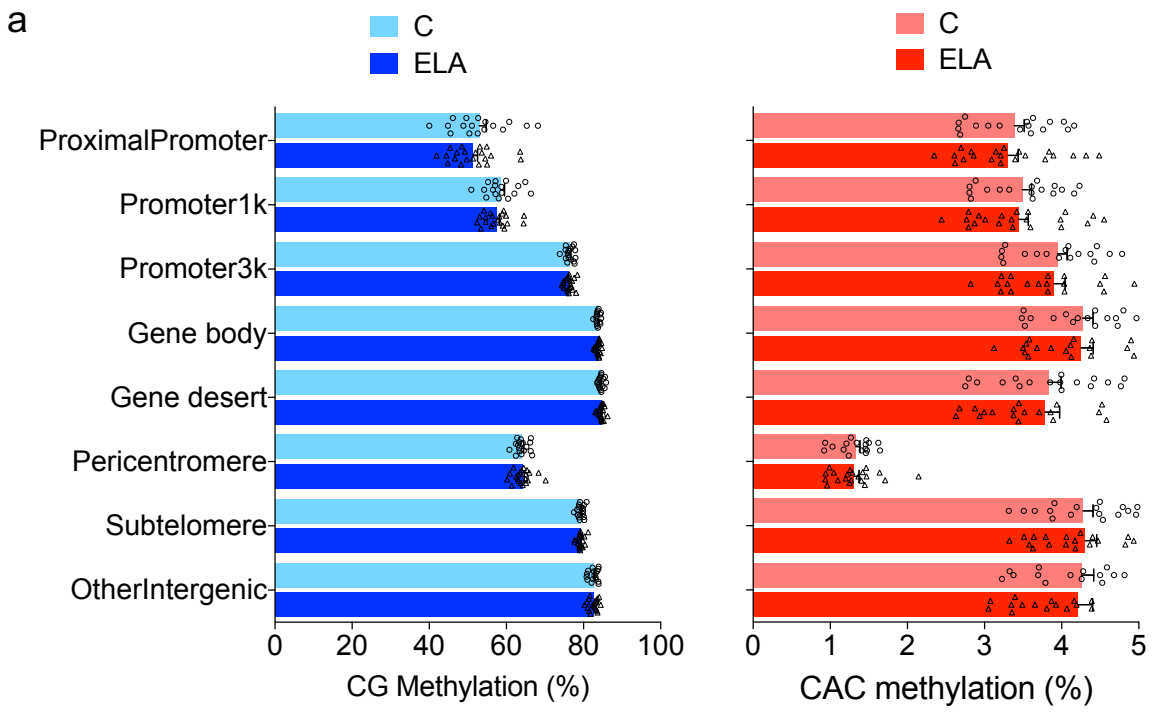
b

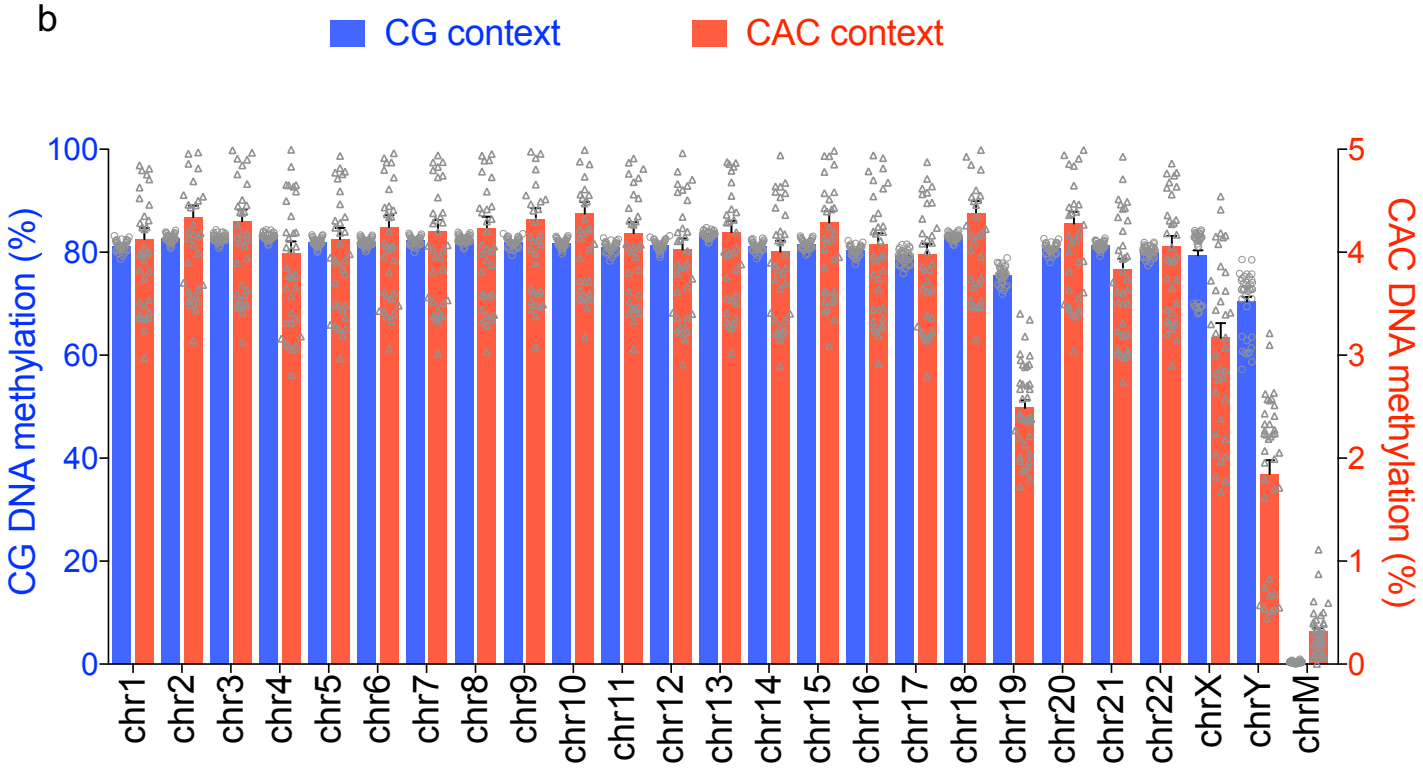


c

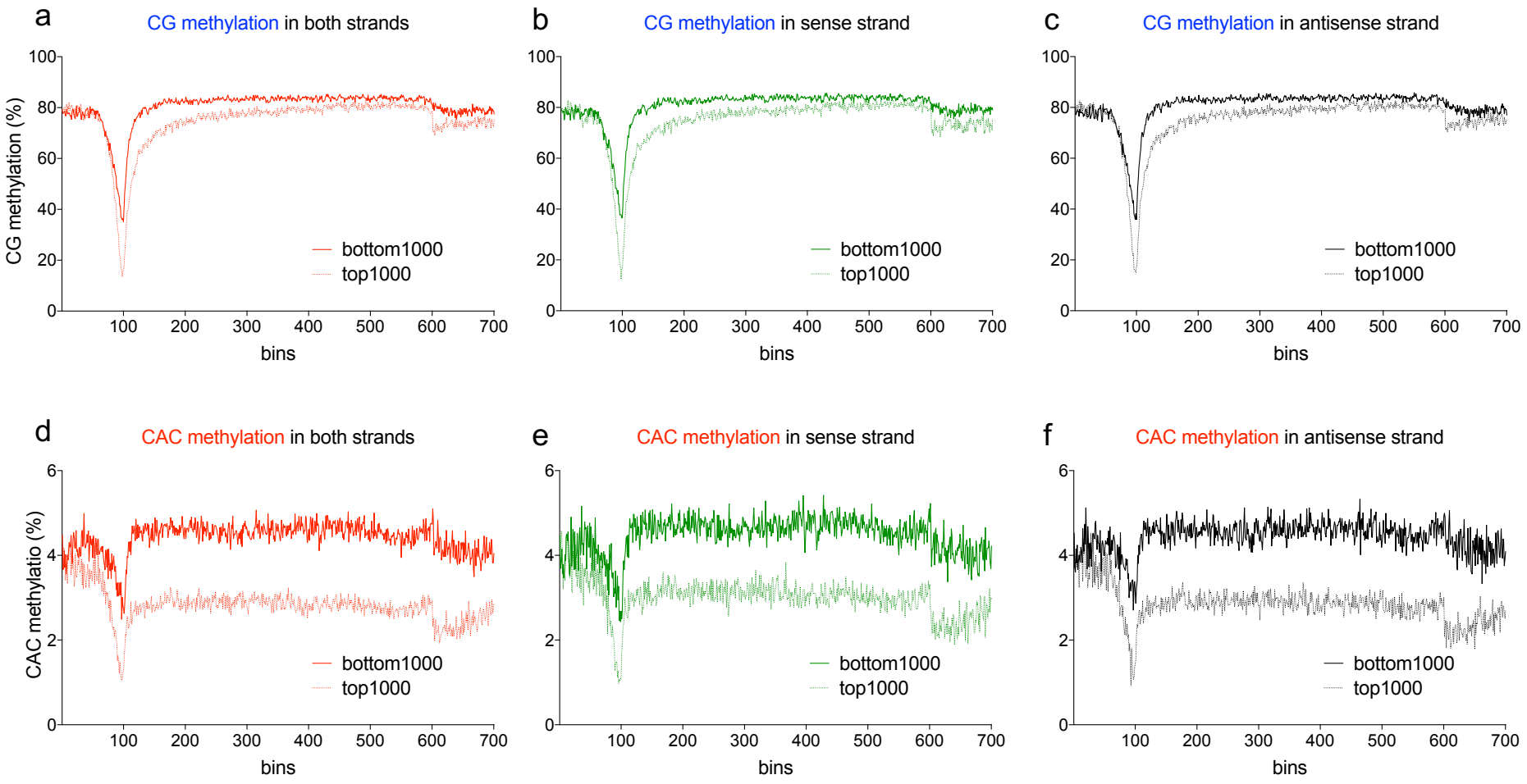


FigureS5

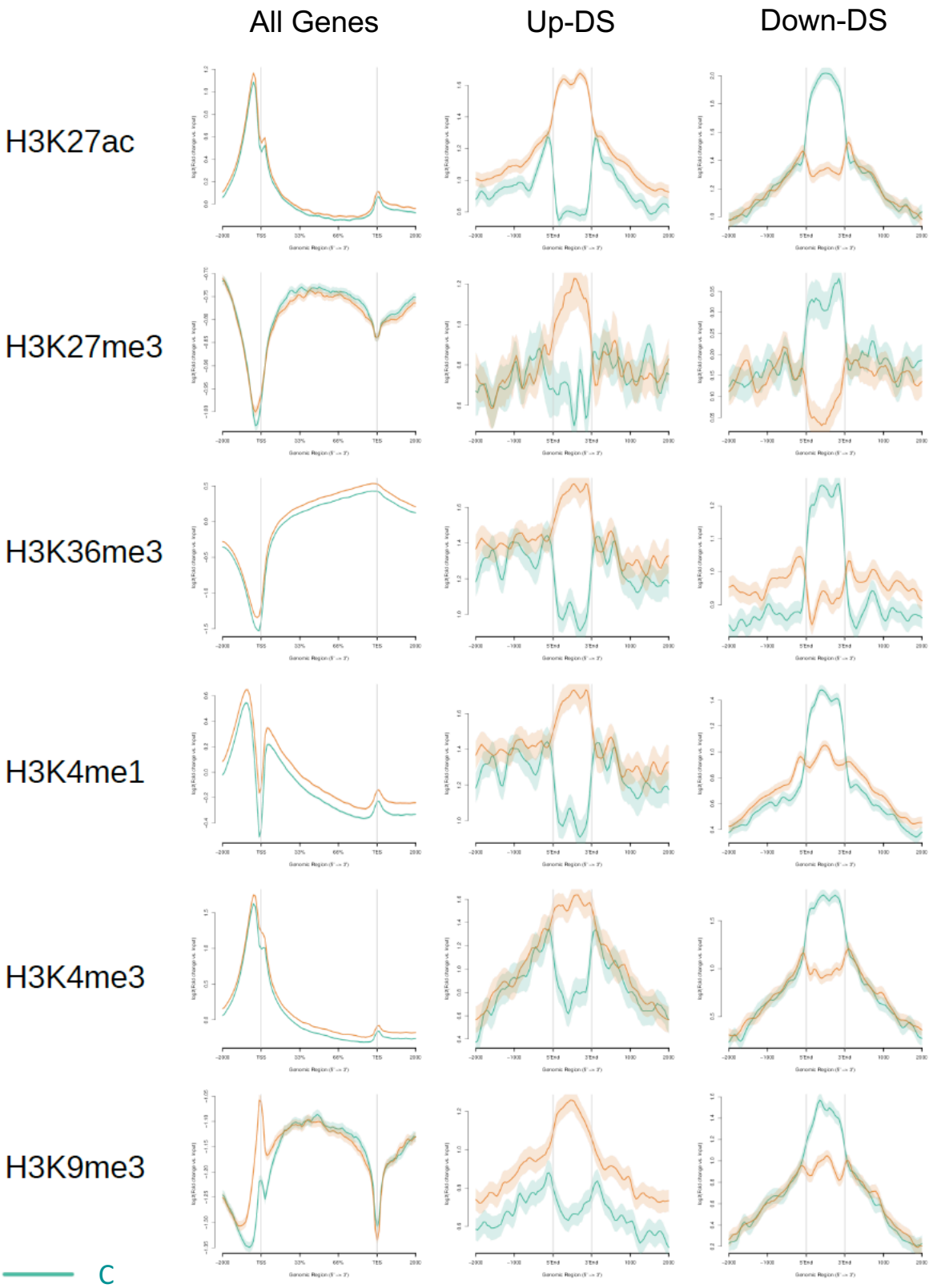

a
C
ELA
C
ELA
ProximalPromoter
Promoter1k
Promoter3k
Gene body
Gene desert
Pericentromere
Subtelomere
OtherIntergenic
0
20
40
60
80
100
CG Methylation (%)
0
1
2
3
4
5
CAC methylation (%)


b
CG context
CAC context
CG DNA methylation (%)
CAC DNA methylation (%)
0
20
40
60
80
100
0
1
2
3
4
5
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
chrM

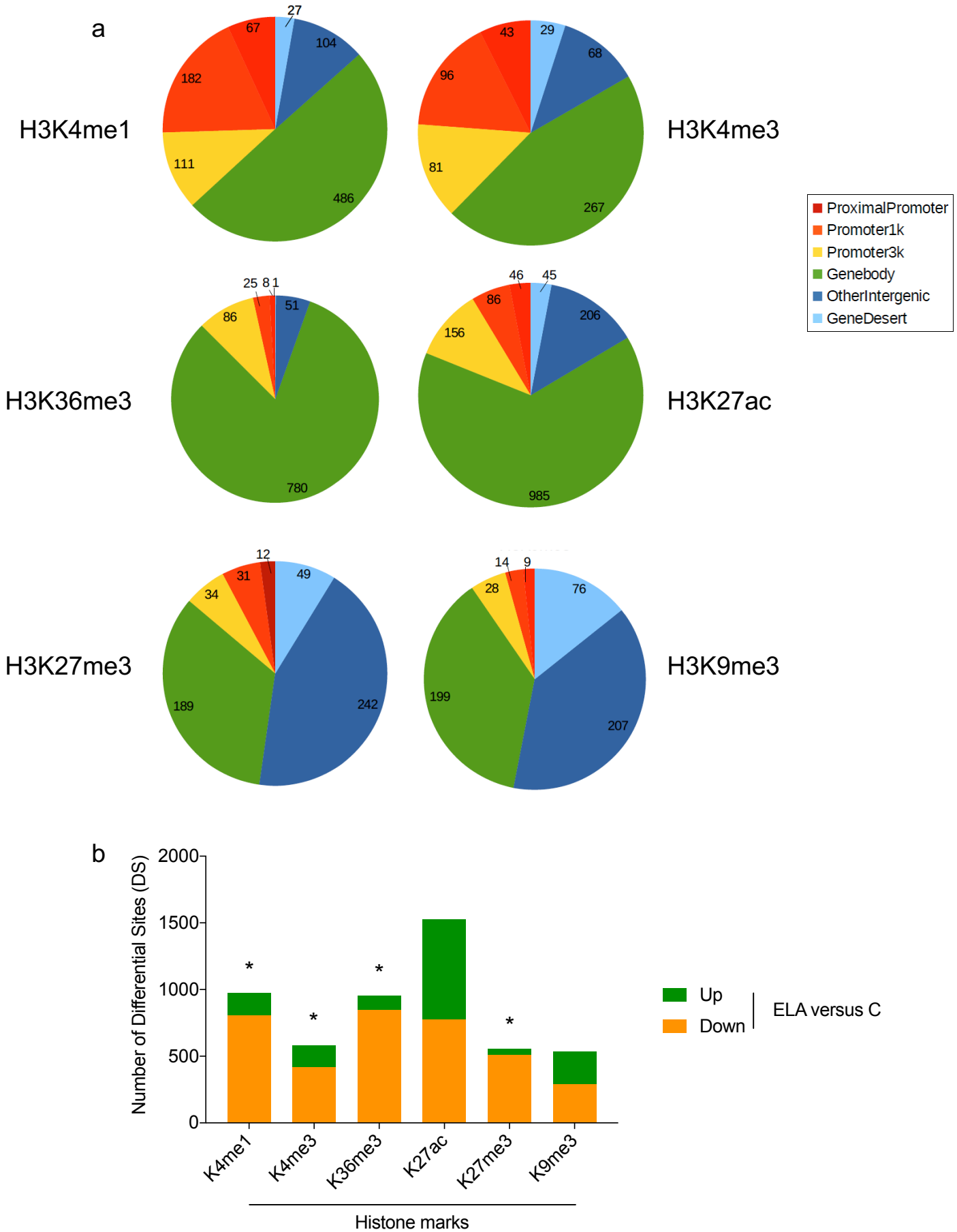
FigureS6


a
CG methylation in both strands
CG methylation (%)
bottom1000
top1000
bins
b
CG methylation in sense strand
bottom1000
top1000
bins
c
CG methylation in antisense strand
bottom1000
top1000
bins
d
CAC methylation in both strands
CAC methylatio (%)
bottom1000
top1000
bins
e
CAC methylation in sense strand
bottom1000
top1000
bins
f
CAC methylation in antisense strand
bottom1000
top1000
bins

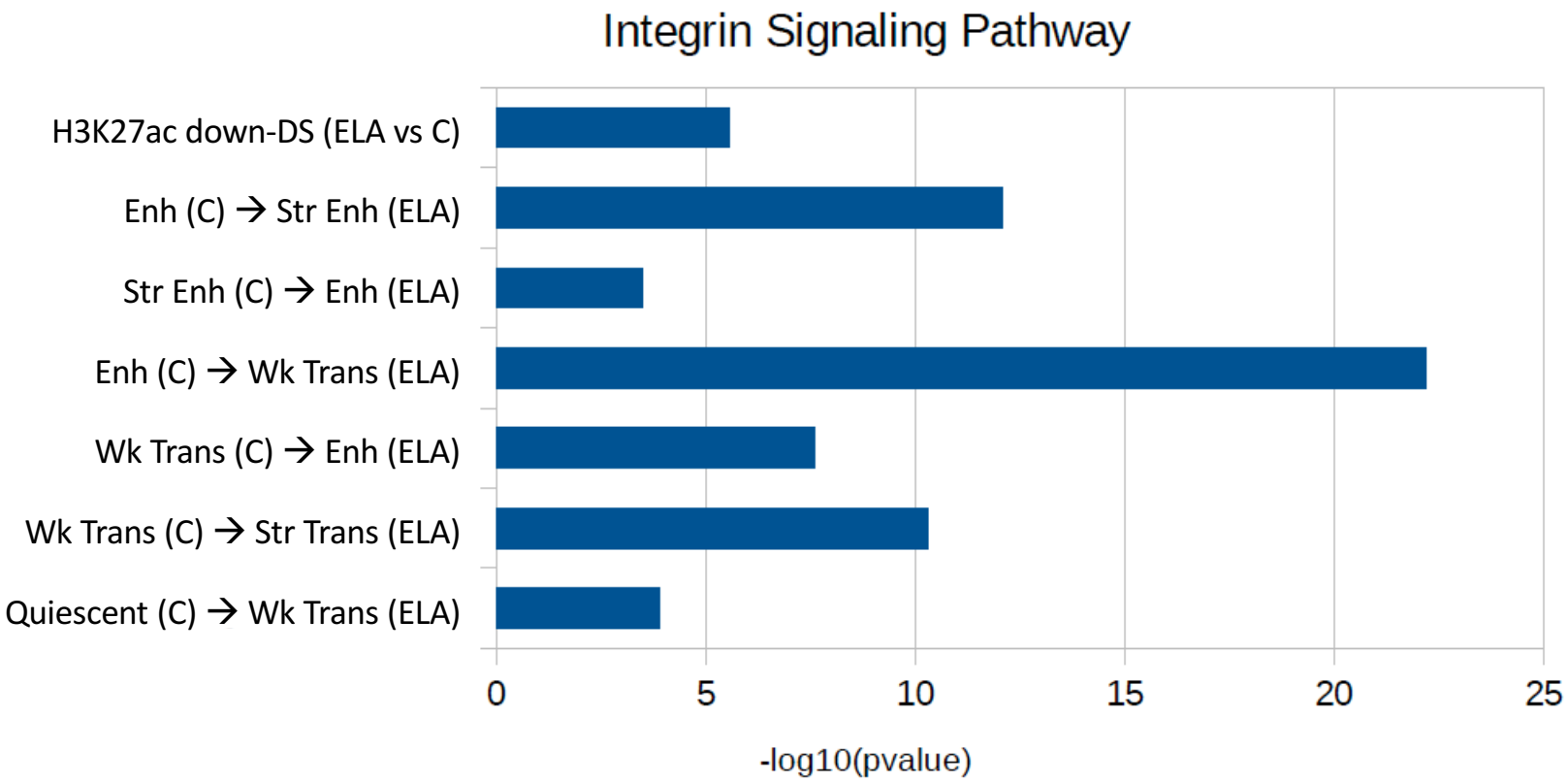
FigureS7


All Genes
Up-DS
Down-DS
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3
C
ELA

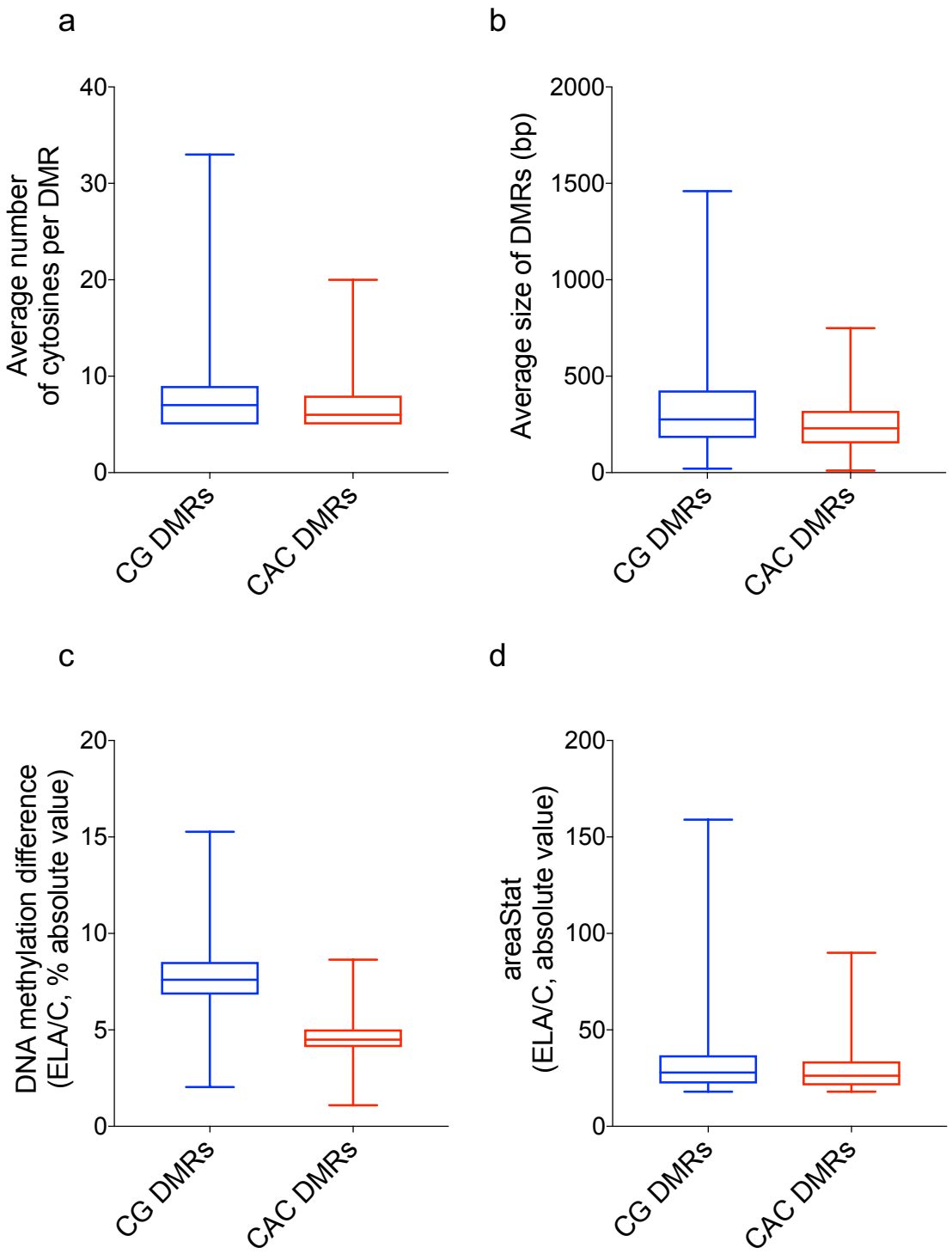
FigureS8


a
H3K4me1
27
67
104
182
111
486
H3K4me3
43
29
68
96
81
267
ProximalPromoter
Promoter1k
Promoter3k
Genebody
OtherIntergenic
GeneDesert
H3K36me3
25
8
1
51
86
780
H3K27ac
46
45
86
206
156
985
H3K27me3
12
31
49
34
189
242
H3K9me3
14
9
28
76
199
207
b
Number of Differential Sites (DS)
2000
1500
1000
500
0
*
*
*
*
K4me1
K4me3
K36me3
K27ac
K27me3
K9me3
Histone marks
Up
Down
ELA versus C

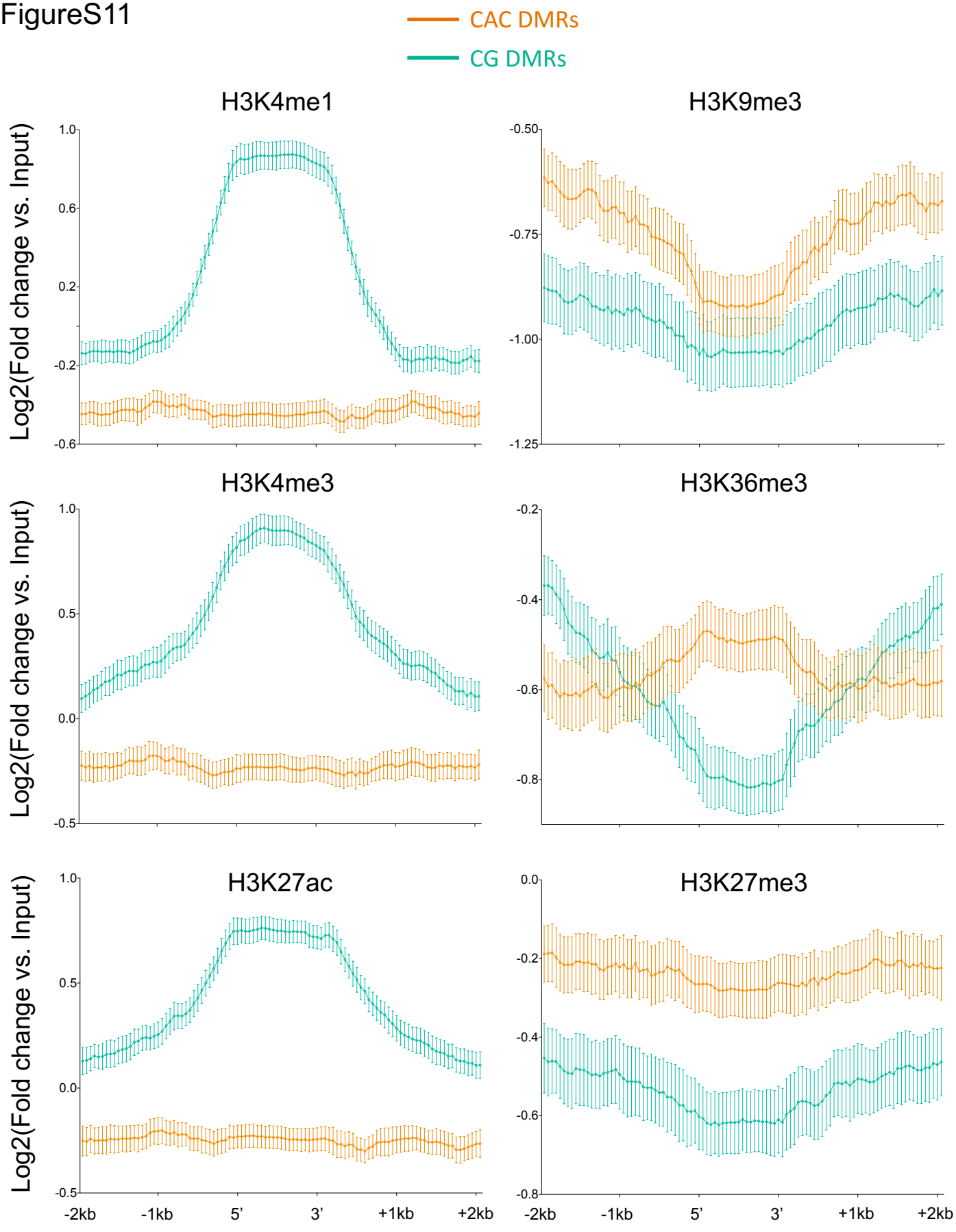
FigureS9


Integrin Signaling Pathway
H3K27ac down-DS (ELA vs C)
Enh (C) → Str Enh (ELA)
Str Enh (C) → Enh (ELA)
Enh (C) → Wk Trans (ELA)
Wk Trans (C) → Enh (ELA)
Wk Trans (C) → Str Trans (ELA)
Quiescent (C) → Wk Trans (ELA)
0
5
10
15
20
25
-log10(pvalue)

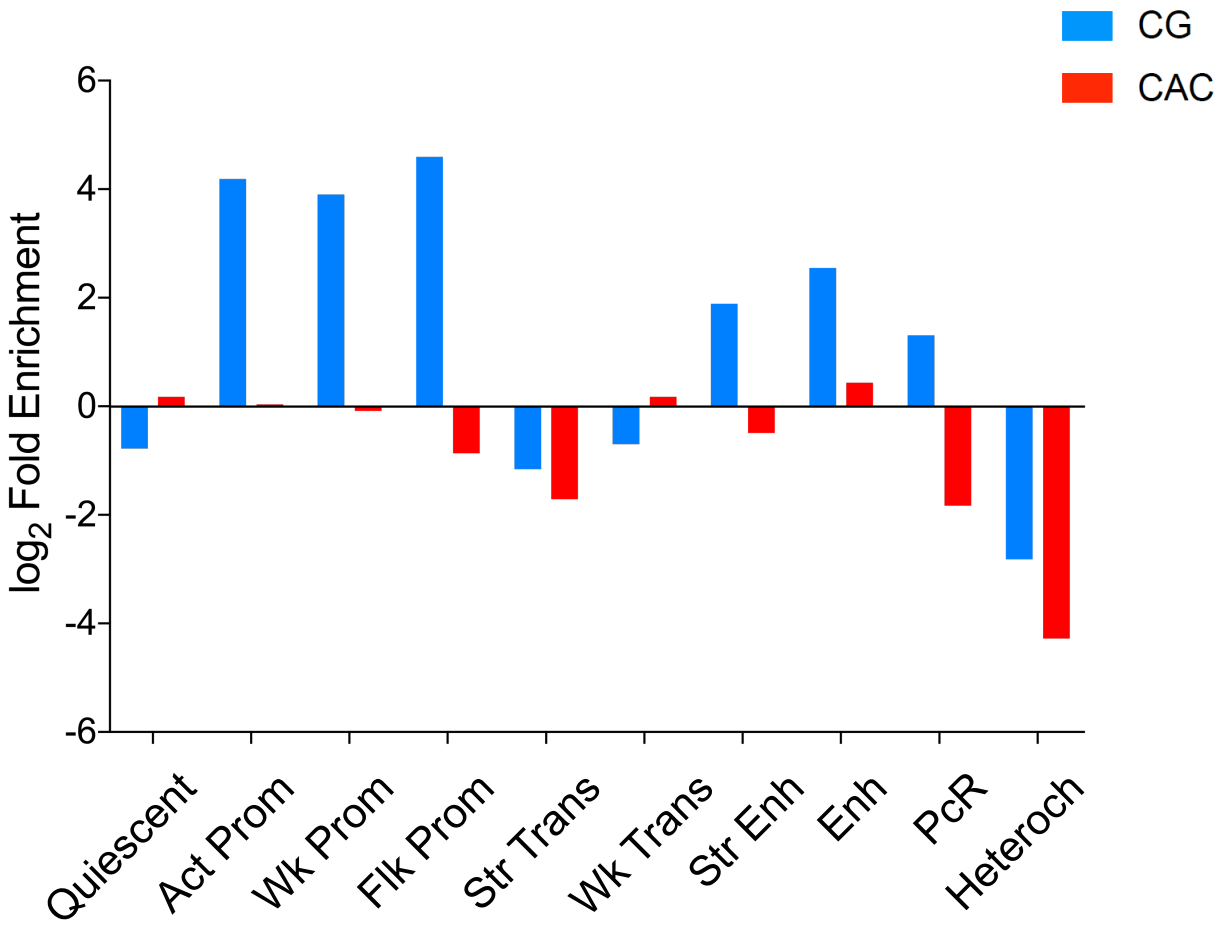
FigureS10


a
Average number of cytosines per DMR
40
30
20
10
0
CG DMRs
CAC DMRs
b
Average size of DMRs (bp)
2000
1500
1000
500
0
CG DMRs
CAC DMRs
c
DNA methylation difference (ELA/C, % absolute value)
20
15
10
5
0
CG DMRs
CAC DMRs
d
areaStat (ELA/C, absolute value)
200
150
100
50
0
CG DMRs
CAC DMRs

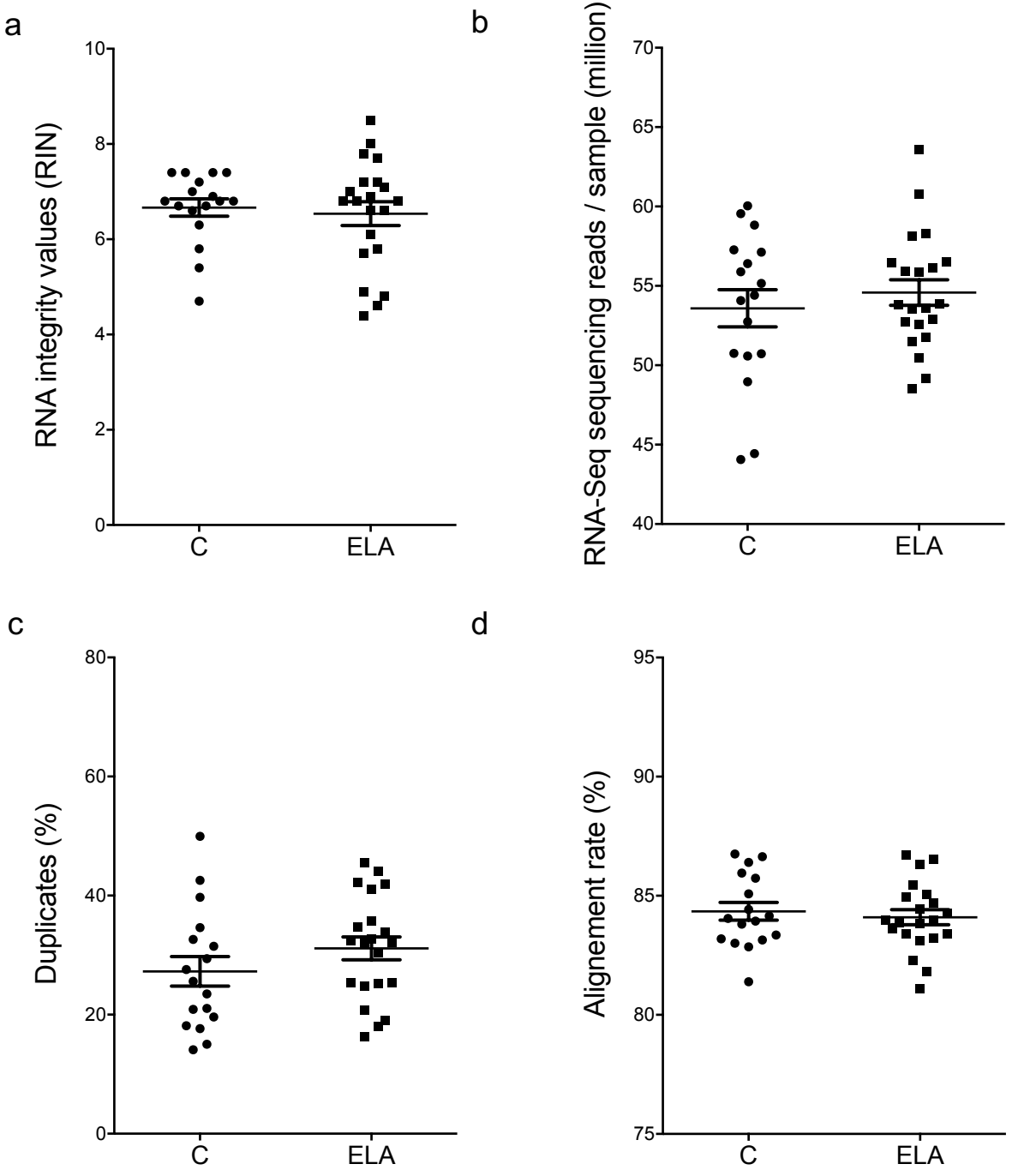
FigureS11


CAC DMRs
CG DMRs
H3K4me1
H3K9me3
H3K4me3
H3K36me3
H3K27ac
H3K27me3
Log2(Fold change vs. Input)
1.0
0.6
0.2
-0.2
-0.6
-0.50
-0.75
-1.00
-1.25
0.5
0.0
-0.5
-0.4
-0.8
-2kb
-1kb
5'
3'
+1kb
+2kb

FigureS12

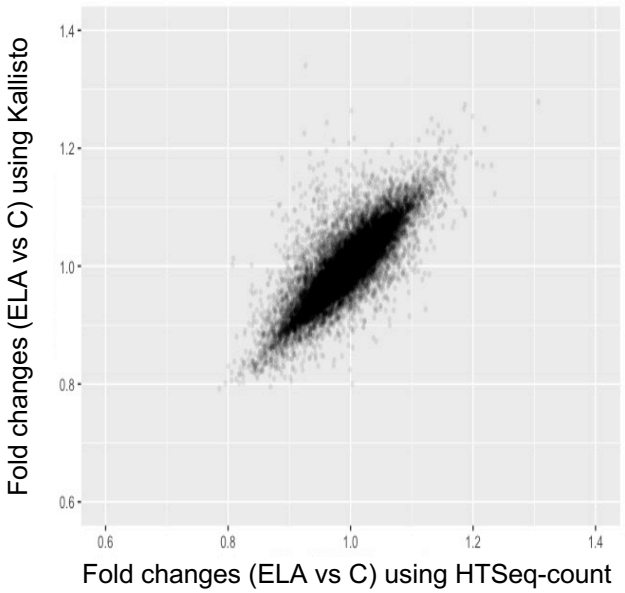

CG
CAC
log2 Fold Enrichment
6
4
2
0
-2
-4
-6
Quiescent
Act Prom
Wk Prom
Flk Prom
Str Trans
Wk Trans
Str Enh
Enh
PcR
Heteroch

FigureS13

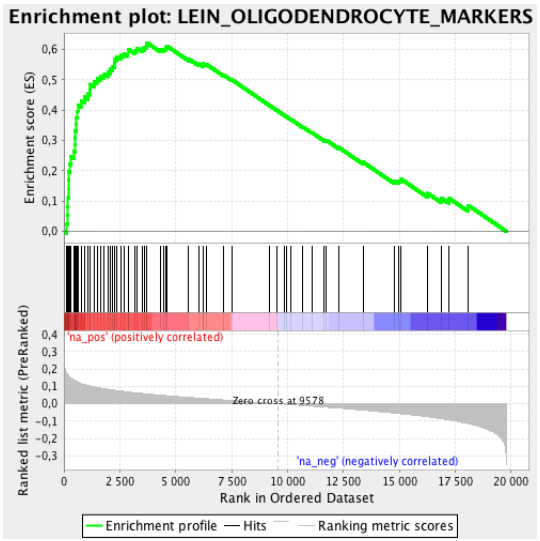

a
RNA integrity values (RIN)
10
8
6
4
2
0
C
ELA
b
RNA-Seq sequencing reads / sample (million)
70
65
60
55
50
45
40
C
ELA
c
Duplicates (%)
80
60
40
20
0
C
ELA
d
Alignement rate (%)
95
90
85
80
75
C
ELA

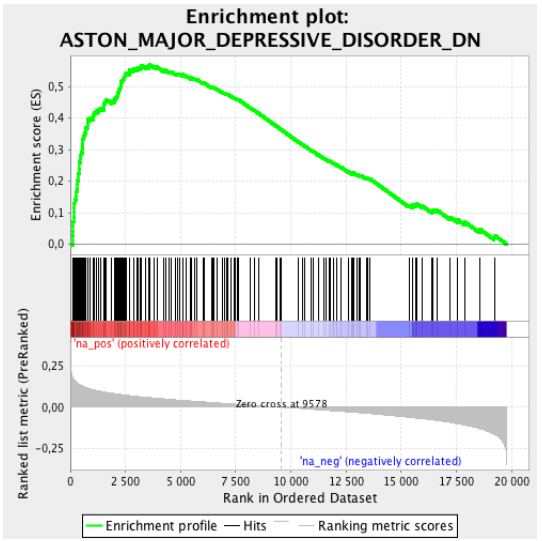
FigureS14

a



b





c